Iniziato	martedì, 7 gennaio 2020, 15:10
Stato	Completato
Terminato	martedì, 7 gennaio 2020, 15:40
Tempo impiegato	30 min.
Punteggio	6,75/15,00
Valutazione	**13,50** su un massimo di 30,00 (**45**%)

Domanda **1**

Parzialmente corretta

Punteggio ottenuto 0,25 su 1,00

For each type of data choose the best suited distance function

High dimensional spaces	Cosine distance ✖
Vector space with real values	Euclidean distance ✔
Boolean data	Manhattan distance ✖
Vectors of terms representing documents	Jaccard coefficient ✖

Domanda **2**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following measure can be used as an alternative to the *Information Gain?*

Scegli un'alternativa:

- a. Rand Index
- b. Silhouette Index
- c. Gini Index ✔
- d. Jaccard Index

Domanda **3**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

How does *pruning* work when generating frequent itemsets?

Scegli un'alternativa:

- a. If an itemset is not frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated ✔
- b. If an itemset is frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated
- c. If an itemset is frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated
- d. If an itemset is not frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated

Domanda **4**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

What measure is maximised by the Expectation Masimisation algirithm for clustering?

Scegli un'alternativa:

- a. The support of a class
- b. The likelihood of an example ✖
- c. The *likelihood* of a class label, given the values of the attributes of the example
- d. The likelihood of an attribute, given the class label

Domanda **5**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

In data preprocessing, which of the following **is not** an objective of the *aggregation* of attributes

Scegli un'alternativa:

- a. Reduce the variability of data ✖
- b. Reduce the number of attributes or objects
- c. Obtain a more detailed description of data
- d. Obtain a less detailed scale

Domanda **6**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following characteristic of data can reduce the effectiveness of K-Means?

Scegli un'alternativa:

- a. Presence of outliers ✔
- b. All the variables are the same range of values
- c. All the variables have the same distribution of values
- d. Presence of values with high frequency

Domanda **7**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

Which of the following is a base hypothesis for a bayesian classifier?

Scegli un'alternativa:

- a. The attributes must have zero correlation
- b. The attributes must be statistically independent inside each class
- c. The attributes must have negative correlation
- d. The attributes must be statistically independent ✖

Domanda **8**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

What is the *cross validation*

Scegli un'alternativa:

- a. A technique to improve the quality of a classifier
- b. A technique to improve the speed of a classifier
- c. A technique to obtain a good estimation of the performance of a classifier when it will be used with data different from the training set
- d. A technique to obtain a good estimation of the performance of a classifier with the training set ✖

Domanda **9**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

What is the meaning of the statement: "*the support is anti-monotone*"?

Scegli un'alternativa:

- a. The support of an itemset never exceeds the support if its subsets
- b. The support of an itemset is always smaller than the support of its supersets
- c. The support of an itemset never exceeds the support if its supersets
- d. The support of an itemset is always smaller than the support of its subsets ✖

Domanda **10**

Parzialmente corretta

Punteggio ottenuto 0,50 su 1,00

After fitting DBSCAN with the default parameter values the results are: 0 clusters, 100% of noise points. Which will be your next trial?

Scegli una o più alternative:

- a. Increase the radius of the neighborhood ✔
- b. Reduce the minimum number of objects in the neighborhood
- c. Decrease the radius of the neighborhood
- d. Reduce the minimum number of objects in the neighborhood and the radius of the neighborhood

Domanda **11**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

In a dataset with D attributes, how many subsets of attributes should be considered for feature selection according to an exhaustive search?

Scegli un'alternativa:

- a. O(D!)
- b. $O(2^D)$ ✔
- c. O(D)
- d. $O(D^2)$

Domanda **12**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

Given the two binary vectors below, which is their similarity according to the Jaccard Coefficient?

a b c d e f g h i j

1 0 0 0 1 0 1 1 0 1

1 0 1 1 1 0 1 0 1 0

Scegli un'alternativa:

- a. 0.1
- b. 0.2
- c. 0.375
- d. 0.5 ✖

Domanda **13**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following *is not* a strength point of *Dbscan* with respect to *K-means*

Scegli un'alternativa:

- a. The *effectiveness*, even in presence of *noise*
- b. The efficiency even in large datasets ✔
- c. The *robustness* with respect to the number of attributes
- d. The *effectiveness* even if there are clusters with non-convex shape

Domanda **14**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the definitions below:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

which of the formulas below computes the *recall* of a binary classifier?

Scegli un'alternativa:

- a. TN / (TN + FP)
- b. TP / (TP + FP)
- c. TP / (TP + FN) ✔
- d. (TP + TN) / (TP + FP + TN + FN)

Domanda **15**

Risposta errata

Punteggio ottenuto 0,00 su 1,00

A Decision Tree is...

Scegli un'alternativa:

- a. A tree-structured plan of tests on multiple attributes to forecast the target
- b. A tree-structured plan of tests on single attributes to obtain the maximum purity of a node ✖
- c. A tree-structured plan of tests on single attributes to forecast the cluster
- d. A tree-structured plan of tests on single attributes to forecast the target

◄ Lab Activity 17-12-2019 - Simulation of lab exa Vai a... Introduction to Big Data - Slides ►